WHAT IS CLAIMED IS:

1. A gene expression state estimating system for estimating the probability of gene expression in each channel, the system including an input device for sending gene expression level data, a program-controlled data analyzer, and an output device, wherein said data analyzer comprises

distributed parameter estimating means for estimating distributed parameters of a mixed normal distribution shown in the following equation (25) using the gene expression level data from said input device, and sending the estimated distributed parameters:

$$(1 - \xi)\,\varphi(u - \mu_0 \mid \sigma_0^2) + \xi\varphi(u - \mu_1 \mid \sigma_1^2) \qquad (25)$$

where $\varphi(* \mid \sigma^2)$ represents the density function of a one-dimensional normal distribution with average 0 and variance σ^2, (μ_0, σ_0^2) and (μ_0, σ_1^2) are average and variance parameters of first and second components, respectively, and ξ is the mixing ratio, with the assumption that $\mu_0 < \mu_1$, $\sigma_0^2 > 0$, $\sigma_1^2 > 0$, $0 < \xi < 1$ is satisfied,

mixing ratio parameter estimating means for estimating a mixing ratio parameter of the mixed normal distribution using the gene expression level data sent from said input device and the distributed parameters sent from said distributed parameter estimating means, and sending the estimated mixing ratio parameter, and

posterior probability calculating means for calculating the posterior probability of the expression state of each gene in each channel using the gene expression level data, the estimated distributed parameters and mixing ratio parameter, and sending the calculated posterior probability to said output device.

2. The system according to claim 1 wherein said distributed parameter estimating means estimates the mixing ratio (ξ), average (μ_0, μ_1), and variance

(σ_0^2, σ_1^2) by applying the mixed normal distribution of two components to data on the sum of the amounts of expression of genes located in a region where the difference of gene expression levels X and Y of two channels is near 0.

3. The system according to claim 2 wherein when the median value of the absolute difference $|v_i|$ $(i = 1, \ldots, n)$ of the gene expression levels X and Y is c_M, the data on the amounts of gene expression is shown by $\{u_i \mid |v_i| \leq c_M,\ i=1, \ldots, n\}$.

4. The system according to claim 3 wherein said distributed parameter estimating means performs estimation by the use of the estimated $\hat{\xi}, \hat{\mu}_0, \hat{\sigma}_0^2, \hat{\mu}_1, \hat{\sigma}_1^2$ according to the following equations (26), (27), (28), and (29) to estimate $\mu, \sigma_\varepsilon^2, \sigma_\beta^2, \lambda$;

$$\hat{\mu} = (\hat{\mu}_1 - \hat{\mu}_0) / 2 \qquad (26)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2\|N_0\|} \sum_{i \in N_0} v_i^2 \qquad (27)$$

$$\hat{\sigma}_\beta^2 = \frac{1}{4} \hat{\sigma}_0^2 - \frac{1}{2} \hat{\sigma}_\varepsilon^2 \qquad (28)$$

$$\hat{\lambda} = \sqrt{\log\left(1 + \frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{4\hat{\mu}^2}\right)} \qquad (29)$$

where N_0 denotes an index set of data values that satisfies $i \in \{i \mid u_i < \hat{\mu}_0\}$ and $\|N_0\|$ denotes the number of elements.

5. The system according to claim 4 wherein said mixing ratio parameter estimating means estimates the mixing ratio parameter $p=(p_{00}, p_{10}, p_{01}, p_{11})$ (where p_{00} denotes a mixing ratio when no gene is being expressed in both cells 1 and 2, p_{11} denotes a mixing ratio when any gene is being expressed in both cells 1 and 2, p_{10} denotes a mixing ratio when a gene is being expressed in cell 1 but not in cell 2, and g_{01} represents a mixing ratio when a gene is being

expressed in cell 2 but not in cell 1) using $\hat{\theta} = (\hat{\mu}, \hat{\lambda}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_\beta^2)$ given from said distributed parameter estimating means by applying a two-variable mixed normal distribution of four components shown in the following equation (30) to the gene expression level data $\{(u_i, v_i)|i=1, \ldots, n\}$ sent from said input device

$$\begin{aligned} & p_{00}g_{00}(u, v \mid \hat{\theta}) + p_{10}g_{10}(u, v \mid \hat{\theta}) + p_{01}g_{01}(u, v \mid \hat{\theta}) + p_{11}g_{11}(u, v \mid \hat{\theta}) \\ & = p_{00}\varphi(u \mid 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2)\varphi(v \mid 2\hat{\sigma}_\varepsilon^2) + p_{10}\varphi_2(u - \hat{\mu}, v - \hat{\mu} \mid \Sigma_{10}) \\ & + p_{01}\varphi_2(u - \hat{\mu}, v + \hat{\mu} \mid \Sigma_{01} + p_{11}\varphi(u - 2\hat{\mu} \mid 4\hat{\mu}^2(e^{\lambda^2} - 1) \\ & + 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2)\varphi(v \mid 2\hat{\sigma}_\varepsilon^2) \end{aligned} \tag{30}$$

where the above equation satisfies the relationships shown in the following equations (31) and (32).

$$\hat{\Sigma}_{10} = \begin{pmatrix} \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2 & \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) \\ \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) & \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 2\hat{\sigma}_\varepsilon^2 \end{pmatrix} \tag{31}$$

$$\hat{\Sigma}_{01} = \begin{pmatrix} \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2 & -\hat{\mu}^2(e^{\hat{\lambda}^2} - 1) \\ -\hat{\mu}^2(e^{\hat{\lambda}^2} - 1) & \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 2\hat{\sigma}_\varepsilon^2 \end{pmatrix} \tag{32}$$

6. The system according to claim 5 wherein said posterior probability calculating means calculates the posterior probability of expression of any gene in cell 1 and cell 2 for each pair (u, v) of the gene expression level data sent from said input device according to the following equation (33) (where $f(u, v \mid \hat{p}, \hat{\theta})$ is given by the following equation (34)) and the following equation (35) (where τ_1, τ_2 take either 1 or 0, which represents the presence or absence of true gene expression in each cell).

$$Pr(\tau_1 = 1 \mid \hat{p}, \hat{\theta}) = \frac{\hat{p}_{10}g_{10}(u, v \mid \hat{\theta}) + \hat{p}_{11}g_{11}(u, v \mid \hat{\theta})}{f(u, v \mid \hat{p}, \hat{\theta})} \tag{33}$$

$$f(u, v \mid \hat{p}, \hat{\theta}) = \sum_{(j,k)\in\{0,1\}^2} \hat{p}_{jk}g_{jk}(u, v \mid \hat{\theta}) \tag{34}$$

$$\Pr(\tau_2 = 1 \mid \hat{p}, \hat{\theta}) = \frac{\hat{p}_{01} g_{01}(u, v \mid \hat{\theta}) + \hat{p}_{11} g_{11}(u, v \mid \hat{\theta})}{f(u, v \mid \hat{p}, \hat{\theta})} \quad (35)$$

7. The system according to claim 5 wherein said posterior probability calculating means calculates the posterior probability indicating an event of differential expression between cell 1 and cell 2 according to the following equation (36)

$$\Pr(\tau_1 \neq \tau_2 \mid \hat{p}, \hat{\theta}) = \frac{\hat{p}_{10} g_{10}(u, v \mid \hat{\theta}) + \hat{p}_{01} g_{01}(u, v \mid \hat{\theta})}{f(u, v \mid \hat{p}, \hat{\theta})} \quad (36)$$

where τ_1, τ_2 take either 1 or 0, which represents the presence or absence of true gene expression in each cell.

8. The system according to claim 6 or 7 wherein said posterior probability calculating means judges whether calculations of posterior probabilities of gene expression have been made for all the pairs (u, v) of the gene expression level data, and when all the calculations have been completed, it ends the process, while when all the calculations have not been completed yet, it calculates the posterior probability related to the next gene, such that the calculated posterior probabilities of gene expression in each channel are sent to said output device, and

said output device displays the posterior probabilities of gene expression in each channel.

9. A gene expression state estimating method for estimating the probability of gene expression in each channel based on gene expression level data, comprising the steps of:

estimating distributed parameters of a mixed normal distribution shown in the following equation (37) using the gene expression level data and sending the estimated distributed parameters:

$$(1 - \xi)\,\varphi\left(u - \mu_0 \mid \sigma_0^2\right) + \xi\varphi\left(u - \mu_1 \mid \sigma_1^2\right) \qquad (37)$$

where $\varphi(* \mid \sigma^2)$ represents the density function of a one-dimensional normal distribution with average 0 and variance σ^2, (μ_0, σ_0^2) and (μ_1, σ_1^2) are average and variance parameters of first and second components, respectively, and ξ is the mixing ratio, with the assumption that $\mu_0 < \mu_1$, $\sigma_0^2 > 0$, $\sigma_1^2 > 0$, $0 < \xi < 1$ is satisfied,

estimating a mixing ratio parameter of the mixed normal distribution using the gene expression level data and the estimated distributed parameters, and sending the estimated mixing ratio parameter, and

calculating the posterior probability of the expression state of each gene in each channel using the gene expression level data, the estimated distributed parameters, and the estimated mixing ratio parameter, and sending the calculated posterior probability.

10. The method according to claim 9 wherein said step of estimating the distributed parameters further comprises a step of estimating the mixing ratio (ξ), average (μ_0, μ_1), and variance (σ_0^2, σ_1^2) by applying the mixed normal distribution of two components to data on the sum of the amounts of expression of genes located in a region where the difference of gene expression levels X and Y of two channels is near 0.

11. The method according to claim 10 wherein when the median value of the absolute difference $|v_i|$ $(i = 1, \ldots, n)$ of the gene expression levels X and Y is c_M, the data on the sum of the amounts of expression of genes is shown by $\{u_i \mid |v_i| \leq c_M,\ i=1, \ldots, n\}$.

12. The method according to claim 11 wherein the estimation is performed in said step of estimating the distributed parameters by the use of the

estimated $\hat{\xi}$, $\hat{\mu}_0$, $\hat{\sigma}_0^2$, $\hat{\mu}_1$, $\hat{\sigma}_1^2$ according to the following equations (38), (39), (40), and (41) to estimate μ, σ_ε^2, σ_β^2, λ;

$$\hat{\mu} = (\hat{\mu}_1 - \hat{\mu}_0) / 2 \qquad (38)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2\|N_0\|} \sum_{i \in N_0} \nu_i^2 \qquad (39)$$

$$\hat{\sigma}_\beta^2 = \frac{1}{4} \hat{\sigma}_0^2 - \frac{1}{2} \hat{\sigma}_\varepsilon^2 \qquad (40)$$

$$\hat{\lambda} = \sqrt{\log\left(1 + \frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{4\hat{\mu}^2}\right)} \qquad (41)$$

where N_0 denotes an index set of data values that satisfies $i \in \{i \mid u_i < \hat{\mu}_0\}$ and $\|N_0\|$ denotes the number of elements.

13. The method according to claim 12 wherein the estimation is performed in said step of estimating the mixing ratio parameter $p=(p_{00}, p_{10}, p_{01}, p_{11})$ (where p_{00} denotes a mixing ratio when no gene is being expressed in both cells 1 and 2, p_{11} denotes a mixing ratio when any gene is being expressed in both cells 1 and 2, p_{10} denotes a mixing ratio when a gene is being expressed in cell 1 but not in cell 2, and g_{01} represents a mixing ratio when a gene is being expressed in cell 2 but not in cell 1) using $\hat{\theta} = (\hat{\mu}, \hat{\lambda}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_\beta^2)$ sent from said step of estimating the distributed parameters by applying a two-variable mixed normal distribution of four components shown in the following equation (42) to the sent gene expression level data $\{(u_i, v_i) \mid i = 1, \ldots, n\}$

$$\begin{aligned} & p_{00}g_{00}(u, v \mid \hat{\theta}) + p_{10}g_{10}(u, v \mid \hat{\theta}) + p_{01}g_{01}(u, v \mid \hat{\theta}) + p_{11}g_{11}(u, v \mid \hat{\theta}) \\ & = p_{00}\varphi(u \mid 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2)\varphi(v \mid 2\hat{\sigma}_\varepsilon^2) + p_{10}\varphi_2(u - \hat{\mu}, v - \hat{\mu} \mid \Sigma_{10}) \\ & + p_{01}\varphi_2(u - \hat{\mu}, v + \hat{\mu} \mid \Sigma_{01} + p_{11}\varphi(u - 2\hat{\mu} \mid 4\hat{\mu}^2(e^{\lambda^2} - 1) \\ & + 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2)\varphi(v \mid 2\hat{\sigma}_\varepsilon^2) \end{aligned} \qquad (42)$$

where the above equation satisfies the relationships shown in the following equations (43) and (44).

$$\hat{\Sigma}_{10} = \begin{pmatrix} \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2 & \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) \\ \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) & \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 2\hat{\sigma}_\varepsilon^2 \end{pmatrix} \quad (43)$$

$$\hat{\Sigma}_{01} = \begin{pmatrix} \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 4\hat{\sigma}_\beta^2 + 2\hat{\sigma}_\varepsilon^2 & -\hat{\mu}^2(e^{\hat{\lambda}^2} - 1) \\ -\hat{\mu}^2(e^{\hat{\lambda}^2} - 1) & \hat{\mu}^2(e^{\hat{\lambda}^2} - 1) + 2\hat{\sigma}_\varepsilon^2 \end{pmatrix} \quad (44)$$

14. The method according to claim 13 wherein the calculation of the posterior probability of expression of any gene in cell 1 and cell 2 is made in said step of calculating the posterior probability for each pair (u, v) of the sent gene expression level data according to the following equation (45) (where $f(u, v \mid \hat{p}, \hat{\theta})$ is given by the following equation (46)) and the following equation (47) (where τ_1, τ_2 take either 1 or 0, which represents the presence or absence of true gene expression in each cell).

$$Pr(\tau_1 = 1 \mid \hat{p}, \hat{\theta}) = \frac{\hat{p}_{10}g_{10}(u, v \mid \hat{\theta}) + \hat{p}_{11}g_{11}(u, v \mid \hat{\theta})}{f(u, v \mid \hat{p}, \hat{\theta})} \quad (45)$$

$$f(u, v \mid \hat{p}, \hat{\theta}) = \sum_{(j,k)\in\{0,1\}^2} \hat{p}_{jk}g_{jk}(u, v \mid \hat{\theta}) \quad (46)$$

$$Pr(\tau_2 = 1 \mid \hat{p}, \hat{\theta}) = \frac{\hat{p}_{01}g_{01}(u, v \mid \hat{\theta}) + \hat{p}_{11}g_{11}(u, v \mid \hat{\theta})}{f(u, v \mid \hat{p}, \hat{\theta})} \quad (47)$$

15. The method according to claim 13 wherein the calculation of the posterior probability indicating an event of differential expression between cell 1 and cell 2 is made in said step of calculating the posterior probability according to the following equation (48)

$$Pr(\tau_1 \neq \tau_2 \mid \hat{p}, \hat{\theta}) = \frac{\hat{p}_{10}g_{10}(u, v \mid \hat{\theta}) + \hat{p}_{01}g_{01}(u, v \mid \hat{\theta})}{f(u, v \mid \hat{p}, \hat{\theta})} \quad (48)$$

where τ_1, τ_2 take either 1 or 0, which represents the presence or absence of true gene expression in each cell.

16. The method according to claim 14 or 15 wherein it is judged in said step of calculating the posterior probability whether calculations of posterior probabilities of gene expression have been made for all the pairs (u, v) of the gene expression level data, and when all the calculations have been completed, the process is ended, while when all the calculations have not been completed yet, the posterior probability related to the next gene is calculated.

17. A gene expression state estimating program for estimating the probability of gene expression in each channel based on gene expression level data, the program instructing a computer to execute the steps of:

estimating distributed parameters of a mixed normal distribution shown in the following equation (49) using the gene expression level data and sending the estimated distributed parameters:

$$(1 - \xi)\,\varphi\left(u - \mu_0 \mid \sigma_0^2\right) + \xi\varphi\left(u - \mu_1 \mid \sigma_1^2\right) \quad (49)$$

where $\varphi(* \mid \sigma^2)$ represents the density function of a one-dimensional normal distribution with average 0 and variance σ^2, (μ_0, σ_0^2) and (μ_1, σ_1^2) are average and variance parameters of first and second components, respectively, and ξ is the mixing ratio, with the assumption that $\mu_0 < \mu_1$, $\sigma_0^2 > 0$, $\sigma_1^2 > 0$, $0 < \xi < 1$ is satisfied,

estimating a mixing ratio parameter of the mixed normal distribution using the gene expression level data and the estimated distributed parameters, and sending the estimated mixing ratio parameter, and

calculating the posterior probability of the expression state of each gene in each channel using the gene expression level data, the estimated distributed parameters, and the estimated mixing ratio parameter, and sending the calculated posterior probability.

18. A computer-readable recording medium storing a gene expression state estimating program for estimating the probability of gene expression in each channel based on gene expression level data, the program instructing a computer to execute the steps of:

estimating distributed parameters of a mixed normal distribution shown in the following equation (50) using the gene expression level data and sending the estimated distributed parameters:

$$(1 - \xi)\,\varphi\left(u - \mu_0 \mid \sigma_0^2\right) + \xi\varphi\left(u - \mu_1 \mid \sigma_1^2\right) \quad (50)$$

where $\varphi(* \mid \sigma^2)$ represents the density function of a one-dimensional normal distribution with average 0 and variance σ^2, (μ_0, σ_0^2) and (μ_1, σ_1^2) are average and variance parameters of first and second components, respectively, and ξ is the mixing ratio, with the assumption that $\mu_0 < \mu_1$, $\sigma_0^2 > 0$, $\sigma_1^2 > 0$, $0 < \xi < 1$ is satisfied,

estimating a mixing ratio parameter of the mixed normal distribution using the gene expression level data and the estimated distributed parameters, and sending the estimated mixing ratio parameter,

calculating the posterior probability of the expression state of each gene in each channel using the gene expression level data, the estimated distributed parameters, and the estimated mixing ratio parameter, and sending the calculated posterior probability, and

outputting the calculated posterior probability.